基于增强特征融合网络的行人再识别 *

邓　滔†，杨　娟，汪荣贵，薛丽霞
(合肥工业大学 计算机与信息学院，合肥 230601)

摘　要： 行人再识别主要是判断不同摄像机捕捉到的行人图像是否属于同一个人。现实生活中，由于人的姿势变化，摄像头的视角变化和背景干扰等因素，导致相同的行人在不同的摄像头产生巨大的差别，这是一项艰巨的任务。近几年，基于深度学习的方法在解决行人再识别问题都取得了显著的效果。然而目前多数方法仅将行人的局部或全局特征分开考虑，从而忽略了行人整体之间的关系，即行人全局特征和局部特征之间的联系。因此，该算法提出了一种增强特征融合网络(Enhanced Feature Convergent Network，EFCN)。在全局分支中，提出适用于获取全局特征的注意力网络作为嵌入特征，嵌入在基础网络模型中以提取行人的全局特征；在局部分支中，提出循环门单元变换网络(Gated Recurrent Unit Change Network，GRU-CN)得到代表性的局部特征，再使用特征融合方法将全局特征和局部特征融合成最终的行人特征，最后借助损失函数训练网络。通过大量的对比实验，该算法网络模型在标准的 Re-ID 数据集上可以获得较好的实验结果。提出的增强特征融合网络能提取辨别性较强的行人特征，该模型能够应用于大场景非重叠多摄像机下的行人再识别问题，具有较高的识别能力和识别精度，且对背景变化的行人图像能提取具有较强的鲁棒性特征。

关键词： 行人再识别；全局特征；局部特征；特征融合
中图分类号： TP391　　doi: 10.19734/j.issn.1001-3695.2020.02.0078

Enhanced feature convergent network for person re-identification

Deng Tao†, Yang Juan, Wang Ronggui, Xue Lixia
(*Dept. of computer & information*, *Hefei University of Technology*, *Hefei* 230601, *China*)

Abstract: Person re-identification is to judge whether the pedestrian across different cameras belongs to the same person or not. While it is challenging task due to the large variations in person pose, occlusion, background clutter, etc. And several deep learning based person re-identification have been proposed and achieved remarkable performance. However, these methods are only considered separately from the local or global features of the pedestrian, ignoring the relationship between the features. So this paper proposed the enhanced feature convergent network (EFCN) . In the global branch, the paper used to employ the new attention to pay close attention to the global feature of pedestrians. In the local branch, it proposed the gated recurrent unit change network(GRU-CN) to obtain more robust local features, and then this paper used feature fusion to connect the extracted global and local features. Extensive comparative experiments show that EFCN can achieve better experimental results on three standard person Re-ID datasets. The proposed enhanced feature convergent network can extract highly discriminative pedestrian features. This model can be applied to the problem of Re-ID under non-overlapping multi-cameras in large scenes. It has high recognition ability and accuracy. The method can extract robust features for pedestrian images with changing background.

Key words: person re-identification; global features; local features; feature convergent

0　引言

行人再识别(person Re-ID)通常是行人检索的子问题，是指在无重叠视域多摄像机监控系统中，辨别两个不同摄像机捕捉到的行人图像是否属于同一个人。person Re-ID 技术可以运用在自动跟踪和检索视频监视网络中的犯罪嫌疑人，能够提高视频监视系统的性能和增加案件处理效率。考虑到行人再识别在视频监控和公共安全中的重要作用，越来越多的研究人员对此问题展开了深入研究。行人再识别主要核心是行人特征表达[1]和特征距离度量。由于监控系统中行人姿势变化，摄像机角度和图片质量问题等因素变化，同一行人在不同的监控摄像头中差异很大，这些问题给行人再识别带来了巨大挑战。具体表现主要如下三个方面：

首先，被捕捉的行人图像在不同的相机中不能对齐，如图 1(a)所示，对于同一个人，在不同图像的相同位置，左侧的红色方框是行人的头部，而右侧的蓝色方框是图像背景。显然，通过卷积神经网络提取的两个区域的特征图存在巨大的差距，无法直接进行比较。为了行人图像解决未对齐问题，文献[2]提出一种基于多特征子空间与核学习的方法，能够有效的识别行人身份信息；文献[3]将关键点直接用于生成感兴趣区域，然后学习行人的局部特征来实现行人的对齐，而这种方法需要训练一个可以达到实际水平的模型，其代价是非常昂贵的。因此本文通过引入注意力嵌入网络去提取行人的全局特征；再使用水平切片方法将提取的全局特征转换为三个相同的局部特征，使得行人特征可以间接对齐，实验效果得到了显著的改善。

其次，如图 1(b)所示，在现实生活中，许多相机拍摄到的行人图像模糊不清，导致图片质量过低，增加了 Re-ID 的

收稿日期： 2020-02-28；**修回日期：** 2020-04-13　**基金项目：** 国家自然科学基金资助项目(61672202)

作者简介： 邓滔(1994-)，男(通信作者)，安徽合肥人，硕士研究生，主要研究方向为深度学习、数字图像处理等(dengtao1994416@qq.com)；杨娟(1983-)，女，安徽合肥人，讲师，博士，主要研究方向为深度学习、智能信息处理等；汪荣贵(1966-)，男，安徽池州人，教授，博导，博士，主要研究方向为深度学习、智能视频处理与分析、视频大数据与云计算；薛丽霞(1976-)，女，四川西昌人，副教授，博士，主要研究方向为数字图像处理、地理信息系统等.

难度。为了解决该问题，文献[4]使用注意力机制关注局部感兴趣区域；文献[5]使用注意力机制从上到下关注局部特征，使用局部特征比较相似性，但是却忽略了全局特征的影响。此外，文献[6]提出了一种多方向显著性学习权值的行人再识别方法,学习到的特征对行人图像具有更好的表述能力，但由于必须将输入图片配对，导致计算效率较低。针对此问题，本文使用注意力机制提取全局特征，并使用水平切片获取局部特征，通过提出的循环门单元变换网络(GRU-CN)，可以着重提取行人的局部重点特征，更好地解决图像模糊问题，还可以减少背景干扰因素。

此外，如图 1(c)所示，当需要区分非常相似的行人图像时，行人细节之间的差异尤为重要。文献[7]通过提取行人的全局和局部特征来捕获行人的细节，但却忽略全局特征与局部特征之间的相关性；文献[8]提出了一种多级相似性度量，通过计算不同级别的相似性得分来识别行人身份，相似性得分的计算量很大。因此，本文在局部分支中提出了循环门单元变换网络(GRU-CN)，该网络可以提取更辨别性的局部特征。同时设计了一种特征融合的方法，将全局特征和局部特征更加紧密地联系在一起，得到了更具代表性的行人特征。本文方法可以更好地提取行人的细节信息，因此对细微差别行人图像的识别效果有明显提高。







(a)行人图片不对齐　(b)图片质量问题　(c)摄像头角度变化及相似行人

图 1　行人再识别的挑战

Fig. 1　Re-ID has some challenges

根据以上分析，在特征学习阶段，提出了一种增强特征融合网络(EFCN)，它具有三个分支：学习全局特征、学习局部特征和学习特征融合。在全局分中，本文把空间注意力和通道注意力相结合作为注意力嵌入式网络 SC-Net，嵌入到 ResNet50[9]网络中；在局部分支中，提出了循环门单元变换网络(GRU-CN)，并使用 GRU-CN 来变换行人局部特征；在特征融合中，利用特征融合操作将全局特征和局部特征融合成新的特征向量。最后把三组特征向量送入损失函数去训练网络参数。

1　模型方法

对于一个给定的查询图像 I_p，行人再识别的目标就是在候选集 G 中去找出与查询图像 I_p 相同身份的其他图像。设候选集 $G=\{I_i\}, i\in[1,\ldots,c]$，其中 c 为行人图像的总数量。让训练集通过网络模型去学习到行人辨识性的特征向量 M 是解决行人再识别的一种方法。在本节中，重点介绍本文提出的增强特征网络模型，如图 2 所示，模型主要分为三个部分，第一部分是全局特征分支，第二部分是局部特征分支，第三部分是特征融合。接下来章节，详细介绍模型的各个部分。

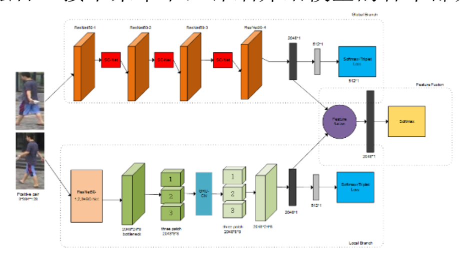


图 2　增强特征融合网络模型的体系结构

Fig. 2　Architecture of the enhanced feature converged network

1.1　全局分支网络

近年来，如何利用深度学习来提取判别特征已受到研究人员越来越多的关注。本文的目的是通过网络模型学习行人的特征图，然后识别行人的身份信息。对于全局分支，利用 ResNet50 作为基础网络。但是由于行人再识别所面临的挑战，如果仅使用基本的 ResNet50 网络来学习全局特征，则提取的全局特征不够代表性，同时引起干扰因素。因此，提出了一种注意力嵌入网络，称为空间和通道注意力嵌入网络(SC-Net)。既是将 SC-Net 与 ResNet50 模型结合起来，其效果能提取出更具代表性的行人全局特征，并稍微修改了 ResNet50 网络，在网络的第四层，删除下采样操作，以获得更大的特征图，其大小为 2048*24 * 8。

接下来介绍嵌入注意力网络的组成，SC-Net 目标是通过注意力机制来增强特征表现力：关注重要的特征，抑制不必要的特征。嵌入注意力网络 SC-Net 主要是由空间注意机制和通道注意力机制组成。由于卷积运算是将跨信道信息和空间信息混合在一起来提取特征的，因此采用该模块来强调通道和空间这两个主要维度的有意义特征。给定行人图像大小为 3*384*128，图像通过 ResNet50 网络得到相应的特征向量。假定特征向量 $F\in R^{C*H*W}$，行人图像通过 ResNet50 第一层得到浅层特征向量 F，其中 C 是通道数， H*W 表示特征向量的长和宽。接下来把特征向量 F 输到 SC-Net 注意力嵌入网络得到特征向量 $f\in R^{C*H*W}$。其具体结构如图 3 所示。

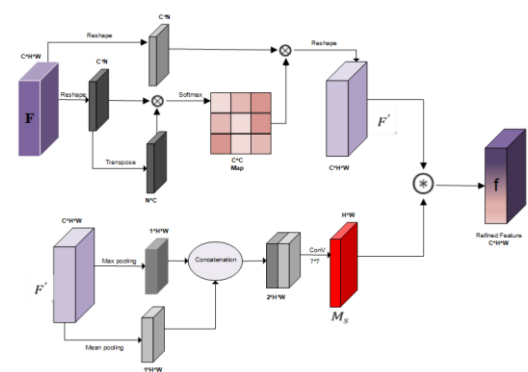


图 3　SC-Net 注意力嵌入网络

Fig. 3　Spatial and Channel attention network

特征向量 $F\in R^{C*H*W}$ 直接地计算通道特征图 $M\in R^{C*C}$，首先，改变特征向量 F 的尺寸大小为 $F_i\in R^{C*N}$，其中 $N=H*W$，F_i 的转置定义为 $F_j\in R^{N*C}$，把 F_i 和 F_j 相乘，最后，利用 softmax 层提取通道注意力特征图 $M\in R^{C*C}$：

$$M_{j,i}=\frac{e^{(F_i\cdot F_j)}}{\sum_{i=1}^{C}e^{(F_i\cdot F_j)}} \tag{1}$$

其中，$M_{j,i}$ 是测量第 i 个通道对第 j 个通道的影响。接下来通道注意机制作用后的特征向量 F' 为

$$F'=\alpha\sum_{i=1}^{C}(M_{j,i}F_i) \tag{2}$$

其中 α 为权重，是从 0 开始学习的，对 M 和 F 的转置进行矩阵相乘。其中 $M_{j,i}\in R^{1*C}$，$F_i\in R^{C*N}$。二者矩阵相乘得到单一通道上的值其大小 R^{1*N}，最后把每个通道值叠加求和得到 $F'\in R^{C*N}$，并将其结果重新定义为 R^{C*H*W}。它有助于提高特征的辨别性。接下来利用通道注意力提取的特征向量 F'，分别采用平均池化和最大池化两种操作得到两组特征向量，接下来将两组特征整合成一个有效的特征描述符。沿着通道方向可以有效地突出重要信息区域。然后，利用一个卷积层作用在特征描述符上，从而得到一个空间注意特征图 $M_s(F')\in R^{H*W}$，式(3)给出计算过程：

$$M_s(F')=\sigma(f([F'_{avg};F'_{max}])) \tag{3}$$

其中: σ 为 sigmoid 函数，f 表示为卷积核为 7*7 的卷积操作，

F'_{avg} 是平均池化得到大小为 $1*H*W$ 的特征向量， F'_{max} 是最大池化得到大小为 $1*H*W$ 的特征向量。综上所述，对于特征向量 $\boldsymbol{F}\in R^{C*H*W}$ ，嵌入注意力网络 SC-Net 是一个通道注意力和空间注意力的组合，其运算过程如下所示。

$$f = M_s(F')\circledast F' \tag{4}$$

其中: $\circledast$ 表示为特征向量的乘积运算; $\boldsymbol{F}'$ 为通过通道注意力优化得到的特征向量，由于嵌入注意力网络 SC-Net 是嵌入学习在 ResNet50 网络的前三个残差块(residual block)后。因此， $\boldsymbol{f}\in R^{C*H*W}$ 为嵌入网络层 SC-Net 优化后的输出特征。接下来通过 ResNet50 的第四层后，获得的特征图为 2048 * 24 * 8。然后利用平均池化获取一个 2048 维特征向量，接下来通过 1*1 卷积层，批归一化和 ReLU 层获得 512 维特征向量。最后，利用三重损失函数和 softmax 损失函数训练了全局分支网络。

1.2　局部分支网络

许多方法主要研究行人的全局特征，会忽略一些行人细节信息，从而加大行人再识别的难度，于是越来越多的研究者考虑行人的局部特征。因此，本文另一个分支是局部分支网络。但是不同于其他方法，本文是利用全局分支提取到的特征向量作为基础。假定在 ResNet50 的第三层中得到一个特征向量$\boldsymbol{F}$；然后利用 bottleneck[9]把特征向量$\boldsymbol{F}$映射为$\boldsymbol{T}$，其尺寸为 2048*24*8；然后利用简单的分块操作，把$\boldsymbol{T}$划分成三块大小为 2048*8*8 相同的特征向量$\boldsymbol{t_1}, \boldsymbol{t_2}, \boldsymbol{t_3}$；最后提出的循环门单元变换网络(gated recurrent unit change network，GRU-CN)变换得到三块局部特征。

对于本文 GRU-CN 网络，其主要结构是在空间转换网络(STN)[10]的基础上作出相关的改进。STN 已经被证明在只有一个前景目标的情况下，可以很好地关注到图像中最重要的部位，还可以自行定位到若干个不同的重要区域。但是通过实验发现 STN 中的归一化网络 Localization net 仅采用简单的卷积操作，不能够满足分块后的局部特征，导致不同小块的局部特征关联性不够强。另外调查研究发现循环门单元(gated recurrent unit，GRU)[11]继承了长短时记忆网络的特点，可以使得局部特征之间更具空间依赖性，同时又提高了计算能力。因此，设计将 GRU 网络放入归一化网络中，并且在其后加入两个全连接层，形成全新的 Localization net。本文提出了循环门单元变换网络 GRU-CN，即能获得更为重要的局部特征信息，又能保持各个局部特征之间的联系性。因此，采用 GRU-CN 可以在局部特征上可以得到行人的重要的区域，然后自动进行特征的对齐，从而得到更好的特征图来提高行人再识别的性能。GRU-CN 网络结构具体见图 4。

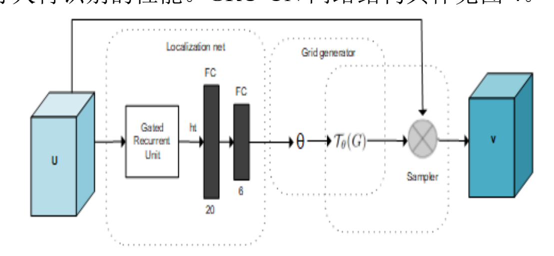


图 4　循环门单元变换网络 GRU-CN 结构

Fig. 4　Gated recurrent unit change network

定位网络的输入是特征图 $\boldsymbol{U}\in R^{C*H*W}$ ，其中 C 为通道数， H 和 W 分别表示高和宽，输出是一个 6 维的仿射变化参数 θ ， θ 可以描述为

$$A_\theta = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix}$$

仿射变换允许输入的缩放、旋转和倾斜。使用定位网络预测转换参数。在网络中，定位网络 Localization net 是循环门单元和两个全连接层的组合，因此

$$(C_t, h_t) = f_{GRU}(U, C_{t-1}, h_{t-1}) \tag{5}$$

$$A_\theta = \mathrm{FC}(h_t) \tag{6}$$

其中: $f_{GRU}(\cdot)$ 为循环门单元， U 是输入特征。$FC(\cdot)$ 为两个全连接层。通过以上的研究，利用 6 维参数 A_θ 来计算图像的仿射变化：

$$\begin{pmatrix} x^s \\ y^s \end{pmatrix} = T_\theta(G) = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix} \begin{pmatrix} x^t \\ y^t \\ 1 \end{pmatrix} \tag{7}$$

其中 $\theta_1,\theta_2,\theta_4,\theta_5$ 是尺寸和旋转参数，而 θ_3,θ_6 是转变参数。式(7)中 (x^t, y^t) 为输出图像的目标坐标， (x^s, y^s) 为输入图像的源坐标。通过 $T_\theta(G)$ 的计算，可以用如下公式得到变化后的特征图 $\boldsymbol{V}\in R^{C*H*W}$ ，其尺寸不变。

$$V_c = \sum_n^H \sum_m^W U_{nm}^c * \max(0, 1-|x^s - m|) * \max(0, 1-|y^s - n|) \tag{8}$$

其中 V_c 表示输出特征图在通道 c 位置上的 (m,n) 上的像素，U_c 表示输入特征图在通道 c 位置上的 (m,n) 上的像素。最后需要导出 U, x^s, y^s ，然后利用损失函数进行反向传播。综上所述，通过利用 GRU-CN 网络的变换，得到更鲁棒性的行人局部信息特征。在局部分支中，同样的通过 1*1 卷积层，批归一化和 Relu 层得到 512 维特征向量。最后，利用三重损失函数和 softmax 损失函数训练局部分支网络。

1.3　特征融合分支

为了更加准确提高行人再识别的识别率，利用特征融合技术把提取到的局部特征和全局特征结合起来以生成更健壮的特征表示是很有必要的。研究发现，将局部特征与全局特征简单地利用 concat 或者 add 操作可能会引入特征噪声干扰。在文献[12]中进行了调查研究，使用向量的外积将使提取的特征图更具表现力。因此，本文使用特征描述符融合操作(feature descriptor fusion, FDF)来融合全局特征和局部特征。为了验证特征描述子融合操作的效果，使用了几组数据进行比较实验，详细的实验在 4.4 节中。假设全局分支提取到的行人特征是 F_g，其中 $F_g\in R^{C*H*W}$，局部分支提取的特征表示为 t_1, t_2, t_3 ,则将三个局部特征利用 concat 得到局部特征向量 F_l ，其中 F_l 的大小与全局特征 F_g 相同。现做以下说明，α_{xy} 表示为全局特征 F_g 中的点(x, y)的特征描述符，而 β_{xy} 被表示为局部特征 F_l 中的点(x, y)的特征描述符。特征描述符融合操作主要使用外部乘积来组合提取的全局特征 F_g 和局部特征 F_l 。 因此，在 F_g 和 F_l 上融合的到融合特征向量 M。具体操作如下：

$$M_{xy} = T(\alpha_{xy} \odot \beta_{xy}) \tag{9}$$

$$\tilde{\mathrm{M}} = \frac{1}{S}\sum_{xy} M_{xy} \tag{10}$$

$$\mathrm{M} = \frac{\tilde{M}}{\tilde{M}_2} \tag{11}$$

其中: $\odot$ 表示向量的外积， $T(\cdot)$ 是把矩阵转换成向量，S 是空间大小，其大小即 $H*W$ 。最终，利用式(11)归一化得到融合特征向量 M 。特征描述符融合操作把局部特征和全局特征融合为 2048 维特征向量。接下来使用 softmax 损失函数来优化特征融合阶段的学习网络参数。在本文网络中，仅在 GRU-CN 网络使用了 dropout[13]策略。最后，将从三个分支提取的特征向量相融合形成行人图像的特征向量。

1.4　损失函数

对于全局分支和局部分支来说，二者共用同一个损失函数。总损失函数是改进的三重损失函数和分类损失函数之和。即

$$\mathrm{Loss} = L_s + loss_{smb} \tag{12}$$

对于 L_s 是分类损失函数，其具体表示为

$$L_s(x_i) = \sum_{i=1}^{T} -y_i \ln f(x_i) \tag{13}$$

其中 y_i 为预测标签， T 为行人的类别， $f(\cdot)$ 为分类函数。而 $loss_{smb}$ 是改进的三元组损失函数[9]，其具体表达式为

$$loss_{smb}(\mu;X)=\sum_{i=1}^{M}\sum_{a=1}^{N}\ln\left(1+e^{\left(d_{m,a,n}^{i,a,p}\right)}\right) \tag{14}$$

$$d_{m,a,n}^{i,a,p}=\max_{p=1\ldots N} Dis\left(y_\mu\left(x_a^i\right),y_\mu\left(x_p^i\right)\right)-\min_{\substack{m=1\ldots M\\ n=1\ldots N\\ m\neq n}} Dis\left(y_\mu\left(x_a^i\right),y_\mu\left(x_n^m\right)\right) \tag{15}$$

其中: M 表示行人的类别，N 表示为每个行人的图像数量，那么 $M*N$ 就表示在一个批次中有 $M*N$ 个三元组。对于 $d_{m,a,n}^{i,a,p}$ 表示为批次中最难三元组损失函数，其中 $\max_{p=1\ldots N} Dis(.)$ 是挑选出正样本中最难的样本；$\min_{p=1\ldots N} Dis(.)$ 则是挑选出负样本对中最难的样本，其中 $Dis(.)$ 是欧式距离函数，也就是说，对于每个样本 x_a^i，其中 x_p^i 是相同行人中对于 x_a^i 的最大距离的图像，x_n^m 是不同行人中最小距离的图像。因此，一个三元组包括 x_a^i,x_p^i,x_n^m，$loss(\mu;X)$ 是一个批次中所有三元组图像的损失函数之和。对于特征融合分支，把融合后的特征向量送入 softmax 分类损失函数中，该损失函数就是式(13)所示。

2　实验结果

在本文实验中，验证模型在行人再识别数据集：Market1501、CUHK03 和 DukeMTMC-reID 上测试。为了使实验简单快捷，所有的实验都是在单查询图像中进行的。本文的网络模型利用 pytorch 深度学习框架，用 NVIDIA GeForce GTX 1080i GPU，Intel i7 CPU 和内存 32GB 训练网络模型。本文采用 adam[14]优化算法来优化模型，adam 是随机梯度下降算法的扩展式。本文随机地把样本分为若干个批次，每批的训练样本的数量为 16 张，每批的测试样本为 16 张。在预处理阶段，对图像进行初始化处理，使得输入图像大小变为 384*128，然后利用数据增强方法。本文利用随机水平翻转和标准化对样本进行增强。在训练阶段，开始先把学习率初始化为 1e-3，然后在 100 个周期后衰减到 1e-4，在 300 个周期后进一步衰减到 1e-5。整个训练过程共达到 400 个周期，共消耗了 8-10 小时使得模型达到拟合状态。

在 Market1501 数据集上评估：在表 1 中展示了 Market1501 数据集的实验结果。将其与近年来最先进的方法进行比较，例如度量学习方法：LOMO＋XQDA[15]，BoW＋KISSME [16]；属性识别学习方法 APR[17]；深度学习方法：GLOB-TO-LOCAL[7]，PCB[18]和 PCB-RPP[18]；使用注意力机制的学习方法：MSCAN [19]，HA-CNN [20]；与当前最新的主力 DMA-CN[21]、Pose[22]算法。从表 1 可以看出，本文方法明显优于度量学习方法。与深度学习方法相比，不需要先验知识，Rank-1 的准确率可以达到 94.4%。与 PCB-RPP 相比，相同的使用注意模型用于辅助学习功能，相比该方法本文的 mAP 增加了 2.2%。本文同样设置一个基本网络模型 baseline，它是以 ResNest-50 网络模型，其 mAP 和 rank-1 达到分别达到 71.59%和 88.84%。从表中可以明显看出，本文方法分别比 MSCAN 和 HA-CNN 的 rank-1 精度高 14.1%和 3.2%。同样，将重排序方法 RK 与本文方法结合使用，可以使 rank-1 达到 95.2%，mAP 达到 93.1%。

在 CUHK03 数据集上评估：CUHK03 数据集是一个具有挑战性的数据集，因为其数据集中存在许多障碍，许多方法均未达到预期的结果。将对实验的结果呈现在表 2 中，并与两组方法进行比较。一方面是低级特征提取方法，另一方面是深度学习方法。在比较方法中，可以清楚地发现，本文方法比低级特征提取方法有显着改进。本文方法在 CUHK03-detected 数据集上的两个评估指标分别为 61.9%和 65.3%。与 PCB＋RPP 方法相比，Rank-1 增加了 2.5%，mAP 增加了 5.2%。在以 CUHK03-label 数据集中，实验准确性分别达到 65.0%和 67.6%。重新排序也与本文方法相结合，实验结果见表 2。

表 1　在 Market1501 数据集上的实验结果

Tab. 1　Experimental results on the Market1501 dataset

网络模型	rank-1	rank-5	rank-10	mAP
LOMO+XQDA	43.7	-	-	22.2
BoW+KISSME	44.4	63.9	72.2	20.8
APR	87.04	95.10	96.42	66.89
GLOB-TO-LOCAL	89.9	-	-	73.9
PCB	92.3	97.2	98.2	77.4
PCB+RPP	93.3	97.4	98.2	80.9
MSCAN	80.3	-	-	57.5
HA-CNN	91.2	-	-	75.7
DMA-CN	88.93	-	-	70.48
Pose	84.3			63.2
baseline(ResNet50)	88.84	-	-	71.59
本文	94.4	98.0	98.6	83.1
本文+RK	95.2	-	-	93.1

表 2　在 CUHK03 数据集上的实验结果

Tab. 2　Experimental results on the CUHK03 dataset

CUHK03-detected	mAP	rank-1	CUH03-label	mAP	rank-1
LOMO+XQDA	11.5	12.8		13.6	14.8
BoW+KISSME	11.7	14.4		12.3	14.2
IDE	19.7	21.3		21.0	22.2
PCB	54.2	61.3		-	-
PCB+RPP	56.7	62.8		-	-
HA-CNN	38.6	41.7-		41.0	44.4
baseline(ResNet50)	59.3	62.1		62.3	64.8
本文	61.9	65.3		65.0	67.6
本文+RK	72.6	73.3		75.4	76.7

在 DukeMTMC-reID 数据集上评估：将本文方法与 DukeMTMC-reID 数据集上已经获得的一些成果的方法进行比较，例如：IDE[1]，ARP，HA-CNN，PCB＋RPP，DMA-CN，Pose。实验结果列于表 3。本文实验效果 mAP 可达到 72.9%，Rank-1 为 86.8%。与更好的 GP-ReID[23]方法相比，尽管 mAP 并没有得到改善，但这可能是由于错误造成的，但是本文的 Rank-1 增加了 1.6%。在此数据集中，本文的方法超越了一些更经典的方法。采用 RK 法，mAP 的实验效果达到 86.8%，Rank-1 达到 89.7%。

表 3　在 DukeMTMC-reID 数据集上的实验结果

Tab. 3　Experimental results on the dukemtmc-reid dataset

网络模型	mAP	rank-1	网络模型	mAP	rank-1
IDE	47.1	67.7	DMA-CN	61.73	78.57
ARP	55.56	73.92	Pose	60.5	78.4
PCB+RPP	69.2	83.3	baseline(ResNet50)	65.1	79.4
HA-CNN	63.8	80.5	本文	72.8	86.8
GP-ReID	72.8	85.2	本文+RK	86.8	89.7

3　实验分析

为了验证本文提出来的各种方法的有效性，接下来在 Market1501 数据集上做了相关的消融实验，利用 baseline 模型去验证各个网络部分的正确性。

3.1　注意力机制网络对实验的影响

对于注意力网络，本文利用四组实验对比证明 SC-Net 的有效性，如表 4 所示.可以明显的发现单独的通道注意力和空间注意力都可以提高实验结果。因为注意力机制被证明能够很好的关注行人的重要信息，减少背景干扰。但是单独的使

用通道注意力和空间注意力可能会导致部分行人信息的丢失，从而降低了识别率。而本文提出的 SC-Net，是从空间和通道两个方面入手，既保持了特征的空间不变性又重点考虑了图像的通道信息，因此，二者结合的注意力机制不仅可以关注重要信息，而且也不会过多的丢失行人图像的信息。综上所述，合并后的嵌入注意力网络 SC-Net 可以有助于提取行人的全局特征，提高识别准确率。

表 4 注意力机制网络的测试结果

Tab. 4 Experimental results of the attention mechanism network

网络模型	mAP	rank-1
baseline(ResNet50)	71.59	88.84
baseline+spatial	74.61	89.31
baseline+channel	73.85	89.10
baseline+SC-Net	78.65	91.34

3.2 局部特征变换网络对实验的影响

通过实验数据表 5 来证明本文的循环门单元变换网络 GRU-CN 的优越性，通过三组对比实验，明显地得出结论，LSTM-STN 或者 STN 网络作用在局部分支上，都可以提高实验的效果；但是本文提出的 GRU-CN，相比较 STN 网络 mAP 提高了 2.12%，Rank-1 提高了 0.52%。对比 LSTM-STN，本文的 GRU-CN 实验结果有微弱的提高。因为相比 LSTM，本文的 GRU 提高了模型的计算能力,如果只是对 GRU 和 LSTM 来说的话，一方面 GRU 的参数更少，因而训练稍快或需要更少的数据来泛化。另一方面，如果你有足够的数据，LSTM 的强大表达能力可能会产生更好的结果。但是行人再识别的数据集的一个缺点就是样本数据的缺少，因此本文的 GRU-CN 比 LSTM-STN 有轻微的提高。通过实验不难发现，STN 网络可以不需要关键点的标定，能够根据分类或者其他任务自适应地将数据进行空间变换和对齐，在输入数据在空间差异较大的情况下，这个网络可以加在现有的卷积网络中，提高分类的准确性。而本文的 GRU-CN 网络一方面保持了 STN 网络的特征对齐性，另一方面使得分块的局部特征产生了联系，更符合行人的整体性信息。因此，GRU-CN 网络作用于局部分支可以提高实验的效果。

表 5 不同变换网络的测试结果

Tab. 5 Experimental results of different transformer network

网络模型	mAP	rank-1
baseline(ResNet50)	71.59	88.84
baseline+STN	73.14	89.66
baseline+LSTM-STN	75.26	90.18
baseline+GRU-CN	75.98	91.69

3.3 分块对实验的影响

本文在局部分支利用水平切分的方法，把提取到的全局特征分为三块局部特征，为了简单的验证本文采用分块数量的有效性，同样地使用几组对比实验来验证分块个数对实验的影响，其实验结果见表 6，从表格中可以明显的得出以下结论。利用分块技术提取行人的局部特征可以提高 baseline 的识别率。但是不同分块数量带来不一样的结果，本文采用的切分成三块的局部特征效果最好。对于切分成两块的局部特征虽然有所提高，但是有可能提取不到行人的细节信息；而采用四块的局部特征会导致行人图像分割过细，引入了背景干扰因素，从而降低了实验结果。因此，利用水平切分成三块的局部特征使得实验能呈现较好的结果。

3.4 特征融合技术对实验的影响

对比基础网络 Baseline,使用全局分支和局部分支的结合，会使得实验结果有明显的提高，实验结果如表 7 所示。同样地为了验证特征描述符融合技术的准确性，本文在表 7 中展示了多组对比实验。表 7 通过与三种融合方法对比：concat 操作、fisher vector(FV)[24]、bilinear[25]。相比简单的 concat 操作，本文方法 FDF 的 mAP 和 Rank-1 分别提高了 2.94%和 3.13%，对比当前主流的融合特征方法 fisher vector(FV)和 bilinear,本文方法 FDF 的实验结果都有轻微的提高。本文的 FDF 方法是从图像的基本点出发，其融合手段保留了特征向量的原始信息，不会像其他融合方法降低了特征的信息。因此，可以得出结论：使用特征描述符融合方法 FDF 有利于提高特征融合后的结果，并且 FDF 方法能够提取更有区分度的行人特征向量。

表 6 分块数量的实验结果

Tab. 6 Experimental results of number of patches

网络模型	mAP	rank-1
baseline(ResNet50)	71.59	88.84
baseline+2 patches	76.37	89.55
baseline+3 patches	78.34	92.41
baseline+4 patches	78.21	91.07

表 7 特征融合技术的实验结果

Tab. 7 Experimental results of feature fusion

网络模型	mAP	rank-1
baseline(ResNet50)	71.59	88.84
global+local	78.1	90.4
global+local+concat	80.06	90.87
global+local+FV	80.98	91.5
global+local+bilinear	81.46	92.43
global+local+本文(FDF)	83.1	94.4

最后，利用三组实验图片，在图 5 中，展示了本文在三个 Re-ID 数据集上的可视化结果，第一列中是待查询图像。根据相似度分数从左到右依次对检索到的图像进行排序。蓝色矩形表示正确匹配的行人，红色矩形表示错误匹配的行人。排序列表的图像是通过本文网络模型获取的。其中 Market1501 和 DukeMTMC-reID 数据集上的识别率都很高；在 Cuhk03 数据集中，在第一组中，查询图像只有五个图像，因此第六个图像不匹配。第二组与行人图片太相似。本文方法产生错误识别，但是在误差范围之内。因此，本文增强特征网络可以较好的识别出行人身份信息。



图 5 在三个数据集上的检索样例

Fig. 5 Sample retrieval results on the three datasets

4 结束语

在本文中，提出了一种增强特征融合网络(EFCN)实现行人再识别。该网络模型可以有效的解决行人再识别行人姿态变化，模糊图像和相似图片等问题。本文将 ResNet50 网络作为基本模型，总体网络模型主要分成三个分支：全局分支，

局部分支和特征融合分支。在全局分支中利用注意力网络 SC-Net 作为嵌入网络，作用在 ResNet50 网络的前三层后，该嵌入网络 SC-Net 与基础网络结合起来能提取到更具表示性的行人全局特征；在局部分支中，主要利用循环门单元变换网络 GRU-CN(gated recurrent unit change network)提取行人重要的局部信息；在特征融合分支，把提取到的全局特征和局部特征利用特征融合方法融合得到鲁棒性和代表性的行人特征，最后利用损失函数训练网络模型。为了验证实验效果，网络模型在三个行人再识别数据集上进行结果评估。通过和不同主流方法相比，其实验效果在三个数据集上都有明显的提高。

参考文献：

[1] 贲晛烨，徐森，王俊. 行人步态的特征表达及识别 [J]. 模式识别与人工智能, 2012, 25 (1): 000071-81. (Ben Xianye, Xu Sen, Wang Jun. Review on pedestrian gait feature expression and recognition [J]. Pattern Recognition and Artificial Intelligence, 2012, 25 (1): 000071-81.)

[2] 齐美彬，颤胜顺，王运侠，等. 基于多特征子空间与核学习的行人再识别 [J]. 自动化学报, 2016, 42 (2): 299-308. (Qi Meibin, Tan Shengshun, Wang Yunxia, *et al.* Multi-feature Subspace and Kernel Learning for Person Re-identification [J]. ACTA AUTOMATICA SINICA, 2016, 42 (2): 299-308.)

[3] Zhao Haiyu, Tian Maoqing, Sun Shuyang, *et al.* Spindle net: person re-identification with human body region guided feature decomposition and fusion [C]// Proc of IEEE the European Conference on Computer Vision (ECCV) . 2017: 907–915.

[4] Zhao Rui, Wanli Ouyang, Wang Xiaogang. Person re-identification by salience matching [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2013: 2528-2535.

[5] Liu Hao, Feng Jiashi, Qi Meibin, *et al.* End-to-End comparative attention networks for person re-identification [C]// Proc of IEEE the 26th Transactions on Image Processing. 2017: 3492–3506.

[6] 陈莹，霍中花. 多方向显著性权值学习的行人再识别 [J]. 中国图像图形学报, 2015, 20 (12): 1674-1683. (Chen Ying, Huo Zhonghua. Person re-identification based on multi-directional saliency metric learning [J]. Journal of Image and Graphics, 2015, 20 (12): 1674-1683.)

[7] Wei Longhui, Zhang Shiliang, Yao Hantao, *et al.* GLAD: Global-local-alignment descriptor for pedestrian retrieval [C]// Proc of the 25th ACM International Conference on Multimedia. 2017: 420–428.

[8] Guo Yiluan, Ngai-Man Cheung. Efficient and deep person re-identification using multi-level similarity [C]// Proc of IEEE on Computer Vision and Pattern Recognition (CVPR) . 2018: 2335–2344.

[9] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2016: 770–778.

[10] Jaderberg M, Simonyan K, and Zisserman A, *et al.* Spatial transformer networks [C]// Proc of the 28th International Conference on Neural Information Processing Systems. 2015: 2017–2025.

[11] Cho K, Van Merrienboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]// Proc of the Conference on Empirical Methods in Natural Language Processing (EMNLP) . 2014: 1724–1734.

[12] Gao Yang, Oscar B, Zhang Ning, *et al.* Compact Bilinear Pooling [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2016: 317–326.

[13] Srivastava N. , Hinton G. Krizhevsky E, *et al.* Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research. 2014, 15 (1): 1929–1958.

[14] Zheng Liang, Yang Yi, Hauptmann A G. Person re-identification: past, present and future [J]. 2016, arXiv preprint: 1610. 02984. https://arxiv.org/abs/1610. 02984.

[15] Liao Shengcai; Hu Yang; Zhu Xiangyu, *et al.* Person re-identification by local maximal occurrence representation and metric learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2015: 2197-2206.

[16] Zheng Liang, Shen Liyue, Lu Tian, *et al.* Scalable person re-identification: a benchmark [C]// Proc of IEEE International Conference on Computer Vision (ICCV) . 2015: 1116-1124.

[17] Lin Yutian, Zheng Liang, Zheng Zhedong, *et al.* Improving person re-identification by attribute and identity learning [J]. pattern recognition 95. 2019: 151–161.

[18] Sun Yifan, Zheng Liang, Yang Yi, *et al.* Beyond part models: Person retrieval with refined part pooling [C]// Proc of the International Conference on European Conference on Computer Vision (ECCV) . 2018: 501–518.

[19] Li Dangwei, Chen Xiaotang, Zhang Zhang, *et al.* Learning deep context-aware features over body and latent parts for person re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2017: 7398–7407.

[20] Li Wei, Zhu Xiatian, Gong Shaogang. Harmonious attention network for person re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2018: 2285–2294.

[21] 刘畅，邱卫根，张立臣. 基于可变形掩膜对齐卷积模型的行人再识别 [J]. 计算机工程与应用, 2020 (03) . http://dx. doi. org/10. 3778. (Liu Chang, Qiu Weigen, Zhang Lichen. Person re-identification based on deformable mask alignment convolution model [J]. Computer Engineering and Applications, 2020 (03) . http://dx. doi. org/10. 3778.)

[22] 裴嘉震，徐曾春，胡平. 融合视点机制与姿态估计的行人再识别方法 [J]. 计算机科学, 2020 (02) . http://dx. doi. org/10. 11896/2019/500013. (Pei Jiazhen, Xu Zengchun, Hu Ping. Person re-identification fusing viewpoint mechanism and pose estimation [J]. Computer Science, 2020 (02) . http://dx. doi. org/10. 11896/2019/500013.)

[23] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-Identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2015: 3908-3916.

[24] Sanchez J, Perronnin F, Mensink T, *et al.* Image classification with the fisher vector: theory and practice [J]. International Journal of Computer Vision. 2013, 105 (3): 222-245.

[25] Lin T, Roychowdhury A, Maji S, *et al.* Bilinear CNN Models for Fine-Grained Visual Recognition [C]// Proc of the international conference on computer vision (ICCV) . 2015: 1449-1457.